IJCSN International Journal of Computer Science and Network, Volume 4, Issue 1, February 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 0.417

A Comparative Analysis of Scheduling Algorithms affecting QoS in Cloud Environment

[1] Nishant Kumar, [2] Mayank Aggarwal , [3] Raj Kumar

[1,2] Department of Computer Science & Engineering, Faculty of Engineering & Technology
Gurukula Kangri University, Haridwar, Uttarakhand-249404, India

[3] Department of Computer Science &Engineering, Faculty of Technology
Gurukula Kangri University, Haridwar, Uttarakhand-249404, India

Abstract - Cloud computing is no longer a buzzword. It has become a common name in the filed of IT and business but there is a lot of scope for better performance and more profit for providers.It deals with several kind of virtualized resources, hence scheduling place an important role in deciding the performace. There are two types of scheduling one for the task and other for allocation of virtual machines. These scheduling schemes affect the Quality of Service of cloud to a great extent. In this paper nine factors are identified affecting QoS and based on these factors exiting algorithms are compared. The result clearly shows that an optimized algorithm for better results in Cloud Computing is needed.

Keywords - *Cloud Computing, QoS, Scheduling Algorithm.*

1. Introduction

Cloud computing refers to the use of computing remotely over network via various ways like as a software, hardware, platform. It is a big turn in Information technology[9]. This architecture is using the resources to its maximum. It's a shift in thinking using hardware for own purpose to using hardware and resources for all over Internet on a paid basis just like renting houses. Here we can rent hardware, platform and also can provide software services on pay per use basis. This model has changed the computing meaning in business. It has changed the way we do business also putting more challenges to researchers to find the best possible algorithm to make things efficient. In short, cloud computing is a convergence of technologies that are making infrastructure and applications more dynamic, highly consumable. It allows organizations to allocate free resources or freed resources rapidly, according to business needs dynamically. Scheduling is one of the most important concepts; It is a mechanisms which controls the order of work to be performed by a computer system. Cloud computing is the service oriented architecture which needs some quality measure to maintain the level of service. This paper mainly focuses on QoS (Quality of Service) affecting factors in previous researches.

1.1 Characteristics

1.1.1 Resource Pooling and Abstraction

In cloud, computing resources are pooled from different systems at one place they seem like one system but at the core there are many system working for that. Multiple consumers can use different physical and virtual resources dynamically assigned according to the demand. So,the cloud has resource pooling and also abstraction that where the work is actually carrying on. The resources includes storage, processing, memory, network bandwidth, virtual machines etc. services.

1.1.2 Rapid Elasticity

Cloud services are scalable. Services can be scaled up in an instant as per the needs.

1.1.3 Custom Pricing

In the architecture like cloud charging consumer for services has various models. 0ne very popular method is pay-per-use. User will be charged only for the services which are used by the user or consumer for a specific time.

1.1.4 Network Sources

Cloud availability is high if network is available. It uses network resources to provide the services.

1.1.5 On-Demand Capabilities

Cloud services are available on-demand basis. If the user needs extra hard-disk space, the user can ask for more

space. If the user need more primary memory for its tasks, user can ask for more ram. Cloud is so flexible that it is available as per user's requirements.

1.2 Service Models

1.2.1 SaaS (Software as a service)

In SaaS model, the consumer uses an application which is deployed in cloud environment, but does not control or know the operating system, hardware or network infrastructure, location which is running the application.

1.2.2 PaaS (Platform as a service)

In PaaS model, the consumer can hire platform (Operating system with minimal requirements) for its applications or other tasks. It controls the applications that run in the environment, but doesn't allow to touch hardware resources over network.

1.2.3 IaaS (Infrastructure as a service)

In IaaS model, the consumer can provide the hardware requirements to the service providers. And service providers will give the user the requested hardware resources and then user can install own operating system and can also deploy own applications.

1.3 Deployment Models

1.3.1 Public Cloud

Public cloud services are characterized as being available to clients from a third party service provider over network or via Internet. The cloud services and resources are procured from very large resource pools that are shared by all end users.

1.3.2 Private Cloud

A private cloud offers many of the benefits of a public cloud computing environment. Private cloud computing systems are like public cloud services offerings within an organization's boundaries to make services accessible for one designated organization. The main advantage of private clouds is that the enterprise retains full control over data, security guidelines, and system performance.

1.3.3 Hybrid Cloud

A hybrid cloud is a combination of a public and private cloud. In this model users typically outsource non-business critical information and processing to the public cloud, while keeping business-critical services and data in their control.

2. Existing Scheduling Algorithms

2.1 An Energy-efficient Scheduling Approach Based on Private Clouds

Cloud computing can lead to large consumption of energy if not properly managed[2].In cloud there are several challenges to effectively schedule virtual machine requests onto computer nodes especially with multiple objectives to meet.[5] covers two such factors,one how to reduce the coming request response time, Second, how to balance the workloads, when the data center is running in low-power mode. And proposes a solution a hybrid energy-efficient scheduling approach which is comprised of pre-power technique and least load first.

To reduce the response time, [5]apply a pre-power technique. When the left capacity of current private cloud is running low, a power-up command would be send to or more asleep nodes to wake up. And when the left capacity overflows the desired spectrum, a power-down command would be send to one or more awake nodes to asleep. Thus, left capacity scales down and more energy would be saved than controlled by the idle threshold.

2.2 Job Scheduling Algorithm Based On Berger Model In Cloud Computing

[1] mainly focused on fairness point of resource allocation and proposed a job scheduling algorithm based on Berger model. The Berger model of distributive justice is based on expectation states. The basic idea of distributive justice is that individual in social system can judge its own gained resources to be fair or not through distribution relationship comparison between itself and other ordinary person in referential structure.[1] classifies user tasks according the QOS parameters.

i. Completion Time
ii. Bandwidth

2.3 Cloud-DLS: Dynamic Trusted Scheduling For Cloud Computing

[12] is inspired by Bayesian Cognitive model. First it proposes Bayesian method based Cognitive trust model then it proposes a trust dynamic level scheduling algorithm named Cloud-DLS. [12] talks about two things Trust and Reliability. Higher the execution ratio of a node, higher the trust, thus increasing reliability. The algorithm efficiently reducing the ratio of task failures.

The basic idea is trust degree defines the level of trust of any node and decided by other recommended action. A computing node displays a message reflecting the characteristics of its behavior when it co-operates with other nodes. A node is duty bound to offer recommendation to other nodes. Thus a node can evaluate its co-partner through its behavior (successful execution ratio). Nodes can also exchange and transmit evaluation messages in order to obtain the trust of target node and guide its co-operation decision. The evaluation of target node's ability of providing services through the reliability shown by its behavior in certain context.

2.4 A Task Scheduling Algorithm Based On QoS-Driven in Cloud Computing

The task scheduling algorithm plays a very important role in deciding the quality of cloud services. [9] covers an efficient algorithm for it. Similarly [13] considers QoS is a major issue which needs to be deal with task scheduling.
Firstly the proposed algorithm [13]computes the priority of task according to the special attributes of task and then sort tasks by priority. The algorithm [13]evaluates the completion time of each task on different service and schedules each task on different services and schedules each task onto a service, which can complete the task as soon as possible according to the sorted task queue.
It gives following qualitative measures for QoS.

a. Completion time
b. Latency
c. Execution price
d. Packet loss rate
e. Throughput
f. Reliability

2.5 Adapting Market-Oriented Scheduling Policies for Cloud Computing

When the user task needs more computational power to complete the task to meet the deadline. Then in such cases federated cloud concept plays good role. If the cloud resources is not enough then current cloud provider can hire services or more computational power from the other cloud providers. To manage federated cloud concept efficiently we need new scheduling policies. [7] proposed two policies or market oriented scheduling policies:

i. Time Optimization Policy

As per algorithm, the scheduler spends the whole available budget for hiring resources from the IaaS providers. After getting accessible, hired resources are added to the list of available serves and the scheduler can dispatch tasks to them. *AddAsServer()* is a thread that keeps track of the request sent to the IaaS provider to get accessible.

ii. Cost Optimization Policy

In each scheduling iteration, a set of tasks are submitted to available resources (local resources). The scheduler estimates the completion time based on the time taken by task that have got completed so far. If the estimation statistics shows that task cannot be completed within time and enough budget is available than one resource can be hired to complete the task.

2.6 Research On Improvement of Task Scheduling Algorithm In Cloud Computing

To carry on many of the sub tasks scheduling at the same time is a complicated problem. By providing services to many users, taking into account the response time of each user, can't keep some users waiting too long. At the same time, the suppliers also want to consider the user's overall satisfaction, so the average completion time of task will be measured. To resolve this issue [6] proposes DFGA algorithm and analyzed to get the best average completion time.

2.7 Multiple Bulk Data Transfers Scheduling among Data Centres

To improve the reliability of the network, [11] proposed MBDTS algorithm to reduce the network congestion. Temporally this algorithm applies store-and-forward transfer mode to reduce the peak traffic load on the link. And spatially, proposed solution is lexicographically minimizes the congestion of all links among datacenters.

2.8 Metaheuristic Scheduling for Cloud

A survey[10] states two problems, efficiency management and utilization of the large amount of computing resources. This paper is using Metaheuristic Algorithm for solving the scheduling problems.In [10] metaheuristic framework states that the number of solutions searched each iteration can be one or more, and the determination operator instead of the selection operator is used to break the limitation by selecting some of the current solutions to pass them on to later iterations. This approach is used to solve scheduling algorithm.

2.9 Batch Task Scheduling-Oriented Optimization Modeling and Simulation in Cloud Manufacturing.

[4] Proposes an algorithm to save time and reduce the cost of workshop production, an optimization model is put first

and then improved particle swarm optimization algorithm used to solve task scheduling problems. While [6] has proposed DFGA for improving task scheduling algorithm.

3. Factors Effecting QoS

3.1 Latency

Latency is the period in which the task execution process is in the queue and waiting for its [5] states two issues in VM scheduling, in which one issue is how to reduce the coming request's response time for that this paper proposed a hybrid energy-efficient scheduling approach which is comprised of pre-power technique. When the left capacity of the current private cloud is running low, in such case a power-up command would be send to one or more asleep nodes to wake up. Therefore the left capacity scales up and coming request would not have to wait that much. Thus lowering down latency. [13] also talks about latency queuing up time. This paper proposed an algorithm which is good enough to reduce the completion time by reducing latency dynamically. [7] states two market oriented policies time and cost optimization policy. In time optimization policy this is consider that the task completion within time is at high priority thus to reduce the request response time it is hiring resources from outside, if the current resources are completely overloaded.

3.2 Reliability

Reliability is the quality of cloud ensuring that all the tasks will be executed; users can rely on cloud services. It is an important factor as large number of users in cloud can lead to starvation. There are scenarios where the work is in the queue and it didn't get the execution time till end.[5] algorithm shows some sign of reliability lowering down latency solves the problem of starvation of tasks, thus increasing execution rate of task and reliability. Further[12] talks about trust which increases the reliability of execution of tasks. Shorten the latency increased the ratio to task completion by removing the problem of starvation and [13] proposed algorithm is dynamically decreasing the latency, thus increasing reliability.

3.3 Energy Consumption

Energy consumption should be less in order to help environment green and low cost execution. High energy consumption means high amount to pay for the task. [5] has got energy efficient algorithm to complete the tasks. It wakes-up nodes only when there is a need. Otherwise the nodes are in sleep mode thus saving energy. This algorithm doesn't use any threshold value to wake up node when workload is high instead it dynamically checks the requirement of the task and cloud scheduler schedules the tasks.

3.4 Completion Time

Completion time is the time to completely finish the task including the queuing time. This should be less. [1] classifies user tasks according the QoS and one is completion time. For real-time systems, the tasks need to be completed within as minimum time as possible. Further paper [7]discussed it in their policy of time optimization that completion of task is more important and for that a cloud can hire more resources and cloud scheduler can scheduled all hired and own resources as per the need in order to fulfill the requirement and minimize completion time.

3.5 Fairness

Fairness is the factor which deals with, that all the resources shall be properly utilized. There shall be no under / over utilization of resources.[1] states that fairness in resource allocation is one another important factor from business and service point of view.

3.6 Bandwidth

The performance of some tasks may depend on BW allocation policy of cloud.So algorithms like [1] are needed which considers BW as a factor for QoS.

3.7 Trust

It is again important factor to think about as there are many service providers are in the market. So to know which is better is important.[12] talks says that trust is another important factor to consider. Trust can get good business. This paper uses novel Bayesian method based cognitive trust model. It gets the trust degree based on node's ratio of successful execution.

3.8 User Privileges

There are many types of service available on cloud. Every user is paying a different amount for different categories and types of service. So this is an important factor to be considered that the user is getting the same for which it is enrolled. Here [13] states that if the priority of a task is high it should be scheduled prior to the lower priority task. Calculation of priority of task has different parameters in which user privileges is also one parameter. User privileges is also mentioned in [7]

where user privileges are measured in business prospects, that if the user task need more computational power, then more resources will be hired for completion of task if user agreed to pay more amount.

3.9 Execution Price

Execution price is high in case user needs its work at high priority. In such scenarios federated cloud are good. Hire services from other cloud also in order to meet the requirement which needs more money.

[7] has discussed two algorithm time and cost optimization. Use optimization as per the requirements. Cost optimization algorithm talks about being economic considering enough time and cost is the important factor which should not exceed. Time optimization algorithm says that if time is more important and cost is flexible the time of execution can't not be compromised.

Table 1:QoS Based Comparison of Scheduling Algorithm

S. No.	*Paper Name*	*Latency*	*Reliability*	*Energy Consumption*	*Completion Time*	*Fairness*	*Bandwidth*	*Trust*	*User Privileges*	*Execution Price*
1.	An Energy-efficient Scheduling Approach Based on Private Clouds [5]	Yes	Yes	Yes	-	-	-	-	-	-
2.	Job Scheduling Algorithm based on Berger model in Cloud environment [1]	-	-	-	Yes	Yes	Yes	-	-	-
3.	Cloud-DLS: Dynamic trusted scheduling for Cloud Computing [12]	-	Yes	-	-	-	-	Yes	-	-
4.	A Task Scheduling Algorithm based on QOS-driven in Cloud Computing[13]	Yes	Yes	-	-	-	-	-	Yes	-

5.	Adapting Market-Oriented Scheduling Policies for Cloud Computing [7]	Yes	-	-	Yes	-	-	-	Yes	Yes
6.	Research on improvement of task scheduling algorithm in cloud computing [6]	-	-	-	Yes	-	-	-	-	-

4. Conclusion & Future Scope

From the above comparison it is evident that the existing algorithms on an average satisfies three of the nine factors considered for QoS i.e one third of the desired.It is impossible to satisfy all the nine factors as they may be complimentary to each other. But work should be done in the area so that we can get more than the current 30% result.

References

[1]. Baominun Xu, Chunyan Zhao, Enzhao Hu, Bin Hu, 2011, "Job Scheduling algorithm based on Berger Model in cloud environment", Advances in Engineering Software 419-425, Elsevier.

[2]. Basmadjian, *2012,* "Cloud computing and its interest in saving energy: the use of a private cloud", 1-5,Journal of Cloud Computing: Advances, Systems and Applications.

[3]. Gao Ming, Hao Li, 2012, "An Improved Algorithm based on Max-Min for cloud task scheduling", Springer.

[4]. Jian, C.F., Wang, Y., Batch task scheduling-oriented optimization modeling and simulation in cloud manufacturing (2014) International Journal of Simulation Modeling, 13 (1), pp. 93-101.

[5]. Jindun Li, Junjie Peng, Zhou Lei, Wu Zhang, 2011, "An Energy-efficient Scheduling Approach Based on Private Clouds", Journal of Information & Computational Sciences 716-724.

[6]. Mingxin, W., Research on improvement of task scheduling algorithm in cloud computing (2015) Applied Mathematics and Information Sciences, 9 (1), pp. 507-516.

[7]. Mohsen Amini Salehi and Rajkumar Buyya, "Adapting Market-Oriented Scheduling Policies for Cloud Computing".

[8]. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, 2009, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems, 25:599_616.

[9]. S. Sindhu and Saraswati Mukherjee, 2011, "Efficient Task Scheduling Algorithm for cloud computing Environment", Springer

[10]. Tsai, C.-W., Rodrigues, J.J.P.C., Metaheuristic scheduling for cloud: A survey (2014) IEEE Systems Journal, 8 (1), art. no. 6516911, pp. 279-291.

[11]. Wang, Y., Su, S., Liu, A.X., Zhang, Z., Multiple bulk data transfers scheduling among datacenters (2014) Computer Networks, 68, pp. 123-137.

[12]. Wei Wang, Guosun Zneg, Daizhong Tang, Jing Yao, 2012, "Cloud DLS: Dynamic trusted scheduling for cloud computing", Expert System with Application 2321-2329, Elsevier.

[13]. Xiaonian Wu, Mengquing Deng, Runlian Zheng, Bing Zeng, Shengyuan Zhou, 2013, "A task scheduling algorithm based on QOS-driven in cloud computing", Procedia Computer Science 1162-1169, Elsevier.